Morphological, and genetic characterization of *Abutilon theophrasti* accessions across a geographic gradient

Teresa Carvalho[1], Eugénia Andrade[1], Isabel M. Calha[1]✉, Dragana Bozic[2], Sava Vrbnicanin[2], José Dorado[3], Jordi Izquierdo[4], Maja Scepanovic[5], Klara Barić[5], Stefano Carlesi[6], Robert Leskovsek[7], Dean Peterson[8], Andrea Veres[9], Vasileios Vasileiadis[10], Roberta Masin[10], Donato Loddo[10]

[1]INIAV I.P. Instituto Nacional de Investigação Agrária e Veterinária, UEIS- Sistemas Agrários e Florestais e Sanidade Vegetal, Quinta do Marquês 2784-505, Oeiras, Portugal
[2]Faculty of Agriculture, University of Belgrade, Nemanjina 6, 11080 Zemun Serbia
[3]Institute of Agricultural Sciences, CSIC, Madrid, Spain
[4] Departament d'Enginyeria Agroalimentària i Biotecnologia, Universitat Politècnica de Catalunya, Castelldefels, Spain
[5]Faculty of Agriculture, University of Zagreb,Croatia
[6]Scuola Superiore Sant'Anna (SSSA),Centro di Ricerche Agro-Ambientali «Enrico Avanzi» (CiRAA)Pisa, Italy
[7] Agricultural institute of Slovenia Hacquetova ulica 17 1000 Ljubljana Slovenia
[8]USDA-ARS Soils Lab, Minnesota, MN, USA
[9] SZIE Plant Protection Institute, H -2103 Godolo, Hungary
[10]Istituto di Biologia Agroambientale e Forestale (IBAF), Consiglio Nazionale delle Ricerche (CNR), Legnaro (PD), Italy
✉ isabel.calha@iniav.pt

Summary: *Abutilon theophrasti* causes -severe yield losses in maize and soybean (70 % - 72%). *A. theophrasti* management is crucial for optimum crop production. Prevention is a pre-requisite to reduce weed pressure and investigation into weed bioecology is needed. This study characterized the morphological and genetic variation of eleven *A. theophrasti* accessions representative of a climatic gradient distribution range from north (47° N) to southeast Europe(39 ° N) and north America (43ºN). To evaluate interpopulation variability, seed were characterized with morphological parameters (length; width; thickness; weight and volume of 100 seeds) and biomolecular markers: trnL-trnF; trnH-trnK; trnT-psbC and matK. Multivariate and univariate analyses were performed to estimate genetic diversity within and among populations. Populations from north and southern Europe were separated in two groups according to seed morphology. However it was not possible to associate seed morphological traits to genetic variability. All gene sequences (cpDNA genome) presented the same nucleotide sequence

Keywords: velvetleaf, interpopulation variability, seed morphology, molecular markers.

1. INTRODUCTION

Abutilon theophrasti Medic (velvetleaf) is an annual plant, belonging to *Malvaceae* family. It is originated from India or China and it was introduced into North America from Europe or Asia by accident, mixed in grains or crop seeds. In the genus *Abutilon*, only *A. theophrasti* occurs in temperate climates. The chromosome number is 2n=6X=42 (Warwick and Black, 1988). Is a self-compatible, autogamous species with approximately 3% of seeds produced in field conditions originatingfrom outcrossing (Andersen, 1988). Propagation is always by seeds, which are produced in large numbers. *Abutilon theophrasti* is among the most important weeds causing significant yield losses in maize and soybean. In absence of weed control these yield losses could reach an average of 70 % and 72% respectively (Campbell and Hartwig, 1982; Sterling and Putnam, 1987). The control of *A. theophrasti*is crucial for an optimum crop production. Prevention is a pre-requisite to reduce weed pressure in any crop and investigation into biological and ecological mechanisms is needed (Northworthy et al*.,* 2012).To accurately infer identities and evolutionary histories of populations it is necessary to use multiple marker systems representing both rapidly evolving and more conserved DNA regions (Soltis and Soltis, 1998). Chloroplast DNA

(cpDNA) sequence variations have been widely used to access interspecific relationships among plants. Despite the low rate of evolution of these sequences, and consequently of their high conservation, Kurokawa and co-authors (2004) used them to distinguish two different haplotypes at the species level. Several authors reported that the non-coding regions have the highest frequency of mutations, as for instance, the *matK* and *trnL-trnF* regions (Taberlet et al*.,* 1991; Nyffeler et al*.,* 2005; Duarte et al., 2011).A collaborative experiment with twelve *A. theophrasti* accessions –representative from a climatic gradient distribution range– was set from 2013-2015 as a larger experiment on seedling emergence variability for European and North American populations (Loddo et al*.,* 2015). In this study, the extent of between- and within-population variation in seed morphological parameters and biomolecular markers of the same *A. theophrasti* populations was investigated. Our hypothesis is that seed morphological traits and seedling emergence could be associated to genetic variability among populations.

2. MATERIAL AND METHODS

2.1. Plant materials. A total of 12 accessions from the species *Abuthilon theofrasti* were used (Table 1). All accessions were obtained from a common experiment (Loddo et al*.,* 2015) being each one from a different country in Europe and USA. These populations are representati-

Table 1. Origin of sampled *A. theofrasti* accessions and morphological characteristics of seeds - average and standard error (SE)

Number	Pop. Code	Country	Site	Lat / Long	Length (mm)	Width (mm)	Thickness (mm)	Weight (g)	Volume (ml)
At 1	CAT	Spain	Lleida	41°37' N	3,72	2,98	1,57	0,92	6,80
				0° 38' E	0,25	0,16	0,17	0,01	0,10
At 2	CRO	Croatia	Čazma	45° 45' N	3,62	2,95	1,55	0,95	6,80
				16° 37' E	0,25	0,15	0,15	0,02	0,10
At 3	GRE	Greece	Makrohori	40° 32' N	3,70	3,00	1,85	0,96	6,80
				22° 14' E	0,25	0,00	0,23	0,01	0,10
At 4	HUN	Hungary	Rackeve	47° 09'N	3,63	2,95	1,52	0,72	6,70
				18° 54' E	0,22	0,15	0,16	0,01	0,10
At 5	IOWA	Usa	Monona	43° 03' N	3,55	2,98	1,63	0,95	6,90
				91° 23' W	0,15	0,16	0,26	0,01	0,10
At 6	VEN	Italy	Legnaro	45° 20' N	3,78	2,98	1,75	0,88	6,70
				11° 58' E	0,25	0,09	0,25	0,01	0,10
At 7	MIN	Usa	Morris	45° 35' N	–	–	–	–	–
				95° 54' W					
At 8	POR	Portugal	Golegà	39° 24' N	3,65	2,97	1,45	0,85	6,50
				8° 29' W	0,27	0,13	0,15	0,01	0,10
At 9	SER	Serbia	RimskiSancevi	45° 40' N	3,67	3,00	1,50	0,89	6,70
				19° 05' E	0,24	0,00	0,13	0,01	0,10
At10	SLO	Slovenia	MurskiČrnci	46° 38' N	3,60	2,98	1,48	0,84	6,90
				16° 06' E	0,31	0,09	0,25	0,01	0,10
At11	CAS	Spain	Arganda del Rey	40° 19' N	3,35	2,98	1,50	0,87	6,70
				3° 29' W	0,23	0,16	0,00	0,01	0,10
At12	TOS	Italy	Pisa	43° 40' N	3,57	2,97	1,80	0,86	6,70
				10° 20' E	0,31	0,13	0,25	0,01	0,10

ve of a climatic gradient distribution range from north (47° N) to southeast Europe(39 ° N) and north America (43ºN). To evaluate interpopulation variability collected seed were characterized with morphological parameters: length; width; thickness; weight and volume of 100 seeds. Thirty seeds were analyzed for each parameter.

2.2. DNA extraction and PCR amplification. DNA was extracted from whole seeds, germinated seeds and seedlings using a CTAB-based method, DNeasy Plant MiniKit (Qiagen, Germany), XpertdirectXtract Kit (Grisp). The quality and the concentration of the DNA were evaluated by ND-2000 Nanodrop spectrophotometer (Nanodrop Technologies). Samples exhibiting the ratios between the absorbance at 260 nm and 280 nm and between 260 nm and 230 nm of 1.7-2.0 and above 2.0, respectively, were directly taken for further analysis. Whenever the values of the ratios were out of the acceptance interval, DNA extracts were purified through columns. Four regions of the chloroplast genome were amplified by using widely used primers (Table 2). The study started with the analysis of trnL/trnF primer pairs.

Table 2. Sources of primers used to amplify three non-coding *loci* from chloroplast genome

Fragment	Primer pairs	Forward and reverse PCR primer sequences (5'-3')	Amplified region
1	trnL/trnF	trnL: GGTTCAAGTCCCTCTATCCC trnF: ATTTGAACTGGTGACACGAG	Intergenic spacer between *trnL*(UAA) 3' exon and the*trnF*.
2	trnH/trnK	trnH: ACGGGAATTGAACCCGCGCA trnK: CCGACTAGTTCCGGGTTCGA	Intergenic spacer between *trnH* and *trnK*non-coding intron
3	trnT-pbsC	trnT: GCCCTTTTAACTCAGTG GTA psbC: GAGCTTGAGAAGCTTCTGGT	
4	matK	MatK2.1:- CCT ATC CAT CTG GAA ATC TTA G- MatK/5R: GTT CTA GCA CAA GAA AGT CG-	*matK* open reading frame*Maturase*
5	rbcL	rbcL_2F: GGA CAT A?? CAA TGC TTT AG rbcL_3R: ATG TCA CCA AAA ACA GAG ACT-	
	rbcL1	rbcLF: ATG TCA CCA CAA ACA GA AAC rbcLR: TCG CAT GTA CCT GCA GTA GC	

PCR reactions were conducted in 25 μL total volume containing 75 ng of total DNA, 2.5 μL of 10x PCR buffer, x μL of 10 mMdNTP mixture, 0.5 μL ofTaq polymerase (Promega) and different concentration of primer pairs (Table 2). Amplification conditions were as follows: 45 cycles of 95 ºC for 10 s, 49 ºC for 30 s and 72 ºC for 1 min. PCR products were observed on an agarose gel (2% m/v) stained with Gel Red and under UV light. After, the fragments were purified according to manufacturer's instructions. The size of the PCR products was assessed by sequencing.

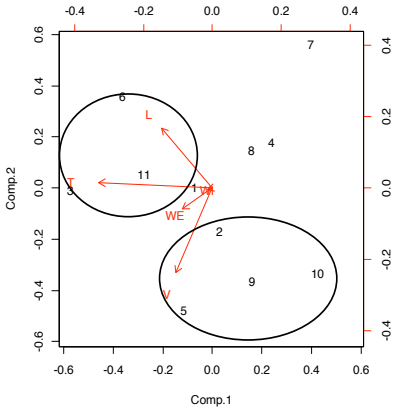
2.3. Sequence analyses. The sequences of the amplified DNA fragments were determined in both directions using the BigDyeTerminator Cycle Sequencing Kit (outsourced service – STABvida, Portugal). A multiple sequence alignment and analysis of the nucleotide sequences were performed using the BioEdit and ClustalW2.1 as a module of Geneious Pro 7.0.5 (Biomatters).

2.4. Sequence alignment and phylogenetic analysis. Sequences were aligned in a multiple sequence alignment and analysis of the nucleotide sequences were performed using the BioEdit and ClustalW2.1 as a module of Geneious Pro 7.0.5 (Biomatters). Bootstrap heuristics

inRAxML (Randomized Axelerated Maximum Likelihood Analysis on the individual data sets (markers) for each accession were performed (Stamatakis et al., 2008). Multivariate and univariate analyses were performed using R (R Development Core Team, 2013) to estimate genetic diversity within populations and among populations.

3. RESULTS AND DISCUSSION

3.1. Seeds morphological characteristics. Only eleven populations were analysed because of seed shortage of population from Minnesota. Seed weight (1000 grains) ranged from 7.2 g to 9.6 g. The mean value 8.81 g of all population studies is in accordance with the Kew database. In general, plant species occupying closed or dry habitats usually produce larger seeds than those found in open or moisture-rich habitats (Zhang, 1997). According to multivariate analysis of our results, populations were classified in two homogeneous groups (Fig. 1): 1 – CAT (1), GRE (3), VEN (6), TOSC (11) with longer and thicker seeds; 2 – CRO (2); IOWA (5); SER (9); CAS (10) with higher volume of seeds. The other three populations could not be classified. The first group includes populations from south Europe, except Portugal (7), and group 2, populations from USA and northern Europe, except Serbia (8) and Hungary (4).Seed weight of *A. theophrasti* seeds varies according to the light conditions of original site. Large seeds were produced by plants in shaded conditions (Zangh et al.,2000). Our study has similar results and it can be attributed to differences in latitude correspondent to lower photoperiod in north (group 2) and longer photoperiod in south Europe (group 1) But the role of light quality (irradiance) and quantity (radiation) are also to be considered. Plant traits are commonly intercorrelated (Schlichting, 1986, 1989) and seed weight has been shown to correlate with other plant traits such as plant height and growth form (Mazer, 1989; Leishman et al., 1995). Thus, it is also worthwhile examining the pattern and degree of integration between seed weight and other traits across populations. Our hypothesis is that seed morphological traits and seedling emergence could be associated to genetic variability among populations.



Importance of components:

	PC1	PC2	PC3	PC4	PC5
Standard deviation	0.1527	0.1174	0.0948	0.0569	0.01427
Proportion of Variance	0.4707	0.2782	0.1815	0.0654	0.00411
Cumulative Proportion	0.4707	0.7489	0.9304	0.9958	1.00000

Figure 1. Multivariated analysis seed parameters of eleven populations of *Abutilon theophrasti.*

3.2. Sequence alignment and phylogenetic analysis. Twelve *A.theofrasti* accessions were considered for this study. For assorted technical reasons, some sequences for some markers still remain incomplete or missing but these sequences will soon be obtained and joined to the total combined plastid DNA data set.

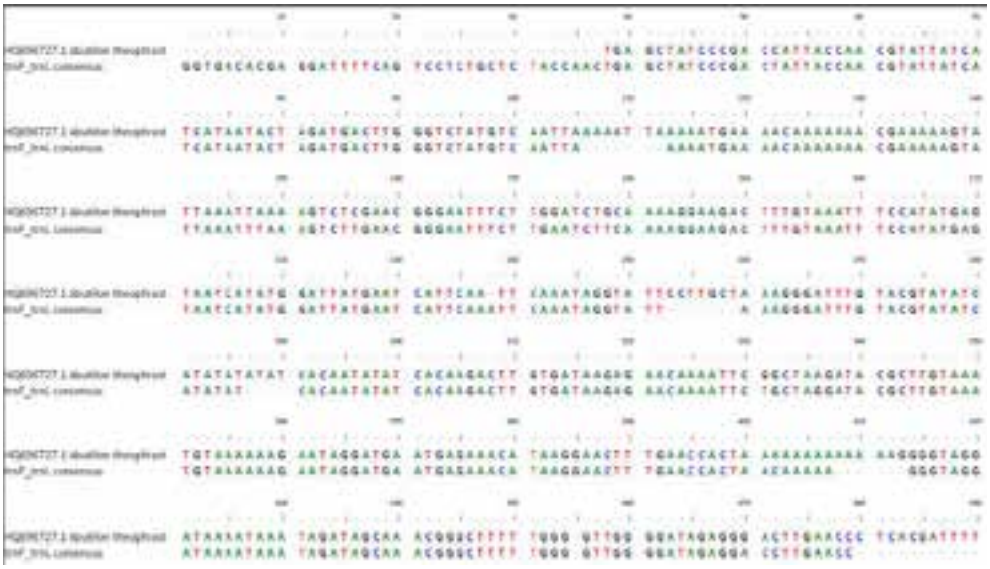


Figure 2. Alignment of consensus trn fragment from Portuguese population (*GenBank MF062945*) with other *Abutilon theophrasti* population.

Previous studies of the genetic relationship between cultivated and weedy *A. theophrasti* and on the phylogenetic analysis of *Eriotheca* and related genera have used a similar technique (Kurokawa et al., 2004; Duarte et al., 2011). Alignment of the *trn*F-*trn*L required no gaps between the 12 accessions.In contrast and surprisingly, when comparing the consensus sequence with those of the NCBI database, the insertion of several gaps was required (Fig. 2). These results were not similar to those of Kurokawa et al. (2004) who studied intra-species variation of 93 *A.theophrasti* weed populations from Japan, USA and Europe: It was possible to classified the weed populations with 'ebony tegument seeds' using chloroplast DNA (cpDNA) into two groups: haplotype A (presence of a one six-base-pair indel) and B (one 30-basepair inversion). Due to the lack of abundance of gene sequences of *Abutilon theofrasti*, we generated these sequences for all the populations used in this study. In our comparison the only variable region was the rbcl that allowd to separate the European populations from the American one (At 5 – IOWA) and the trnF-trnL region that showed to be different from the *Abutilon theofrasti* population used in the study conducted by Duarte et al (2011). All the sequences obtained, for all the other markers, were aligned. The European populations were equal to each other and to a Portuguese population originated from Azores (Sxchaefer et al., 2011). In our study the mother effect was

not strongly expressed as seeds from all populations came from a common experiment.The plastid is generally uniparentally inherited (Birky, 2001) and behaves as a single non-recombining *locus*, providing a strong signal of population and phylogenetic history (Petit and Vendramin, 2007) The amount of variation present over short regions may be too low to distinguish recently diverged taxa (Piredda et al*.,* 2011). One reason for difficulty in differentiate populations using chloroplast genome could be a recent and common origin of European and American populations, compared to ancient Asiatic populations where this species is originated. *A.theophrasti*was recently introduced in Europe. In the Iberian Peninsula there are reports from the XIX century but it became a serious weed in summer crops in 1980's (Recasens et al*.,* 2003).

4. CONCLUSIONS

Populations from north and southern Europe were separated in two groups according to morphologicalcharacteristics of *Abutilon theophrasti* seeds which, as a hypothesis could be attributed to differences in latitude, concerning light quality and quantity. Among the European populations was not observed variability, however, all showed to be different from a population from North America. Additionally, all the populations analyzed in this study showed to be different from a South American population. The preliminary results in this study did not allowed for a possible association between seed morphological traits and genetic variability among populations of *A.theophrasti*. One of the reasons may be the recent origin of the populations of *A. theophrast* as weeds in Europe. Still, they are indicative of higher levels of variation among the populations of different continents. AS few non-coding cpDNA regisons were surveyed, a larger number of genetic markers of the chloroplast and nucleus are needed to increase the probability to detect differences among populations.

5. REFERENCES

Andersen RN, 1988. *Weed Science,* 36(5), 599-602.
Campbell RT, Hartwig NL, 1982. *Proc. Northeastern Weed Science Society*, 36, 2-4.
Kurokawa et al (2004). *Heredity*, 93, 603-609.
Loddo et al (2015). *17th EWRS «Weed Manag Changing Environments»*, 23-26 June, Montpellier, F.
McCauley DE (1995). *Trends in Ecology and Evolution*, 10, 198-202.
Norsworthy et al. (2012). *WeedSience*, 60, 31-62.
R Development Core Team (2013). *R: A Language and Environmt for Statistical Computing*. Vienna.
Recasens et al. (2003). *Contribuciones al Conocimiento de las Especies Exóticas Invasoras. Grupo Especies Invasoras* Ed., G.E.I. Serie Técnica, 1, 213-215.
Soltis DE, Soltis PS (1998) *DNA Sequencing:* 1-41. Kluwer Academic Publishers, Boston, MA, USA.
Sterling TM, Putnam AR (1987*). Weed Science*, 35(3), 308-314.
Warwick SI, Black L D (1988). *Canadian Journal of Plant Science,* 68, 1069-1085.
Zhang et al. (1997). *Canadian Journal of Botany*, 75(9), 1614-1620.